Inference for likelihood-based estimators of generalized long-memory processes

Beaumont, Paul and Smallwood, Aaron

Florida State University, University of Texas-Arlington

30 September 2019

Inference for likelihood-based estimators of generalized long-memory processes☆

Paul M. Beaumont[a], Aaron D. Smallwood[b,*]

[a] *Department of Economics, Florida State University, Tallahassee, FL 32306, USA*
[b] *Department of Economics, University of Texas Arlington, 701 S. West Street. (mailbox: 19479), Arlington, TX 76019, USA*

Abstract

Despite a recent proliferation of research using cyclical long memory, surprisingly little is known regarding the asymptotic properties of likelihood-based methods. Estimators have been studied in both the time and frequency domains for the Gegenbauer autoregressive moving average process (GARMA). However, a full set of asymptotic results for all parameters has only been proposed by Chung (1996a,b), who present somewhat tenuous results without an initial consistency proof. In this paper, we review the GARMA process and the properties of frequency and time domain likelihood-based estimators using Monte Carlo analysis. The results demonstrate the strong efficacy of both estimators and generally support the proposed theory of Chung for the parameter governing the cycle length. Important caveats await. The results show that asymptotic confidence bands can be unreliable in very small samples under weak long memory, and the distribution theory under the null of an infinitely long cycle appears to be unusable. Possible solutions are proposed, including the use of narrower confidence bands and the application of theory under the alternative of finite cycles.

Keywords: long memory, GARMA, CSS estimator, Whittle estimator
JEL Classification Codes: C22, C40, C58, G12

☆The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that were used to generate simulation results in the paper. URL: http://www.tacc.utexas.edu.

*Corresponding author: Aaron D. Smallwood, Department of Economics, University of Texas Arlington, 701 S. West Street (mailbox: 19479), Arlington, TX 76019, USA

Email addresses: beaumont@fsu.edu (Paul M. Beaumont), smallwood@uta.edu (Aaron D. Smallwood)

1. Introduction

Few contributions to time series analysis have fomented more interest than the introduction of long memory by Granger and Joyeux (1980) and Hosking (1981). These methods allow for slowly decaying autocorrelation functions and the existence of spectral density functions with one or more singularity. In economics, long memory has provided a major breakthrough in allowing researchers to bridge the gap between unit roots and transitory I(0) dynamics.

As emphasized by Dissanayake et al. (2018), attention has recently focused on methods that can accommodate cyclical long memory, including seasonal long memory and Gegenbauer autoregressive moving average (GARMA) models. The GARMA model, which has received specific attention, is defined as follows,

$$(1 - 2\eta L + L^2)^\lambda \phi(L)(x_t - \mu) = \theta(L)\varepsilon_t \tag{1}$$

where $\phi(L)$ and $\theta(L)$ are p and q order polynomials in the lag operator L, and ε_t is a mean zero disturbance sequence with $E(\varepsilon_t^2) = \sigma^2$ and no serial correlation. With $\nu = \cos^{-1}(\eta)$ the process possesses a spectral density function given by

$$f(\omega) = \frac{\sigma^2}{2\pi} \left| \frac{\theta(e^{-i\omega})}{\phi(e^{-i\omega})} \right| 2|\cos(\omega) - \cos(\nu)|^{-2\lambda}. \tag{2}$$

When $p = q = 0$, the process has an autocorrelation function at lag j that is proportional to $\cos(j\nu)j^{2\lambda-1}$. When $\nu = 0$, the result is an $ARFIMA(p, 2\lambda, q)$ process as originally studied by Granger and Joyeux (1980) and Hosking (1981). The process above is covariance stationary provided $\lambda < 1/4$ when $\nu \in \{0, \pi\}$ or when $\lambda < 1/2$ otherwise (Gray et al. 1989).

This model, and its extension, the k-factor GARMA model, have been applied across virtually every discipline that uses time series methods. Examples include atmosphere C02 (Woodward et al. 1998), sunspots (Chung 1996b; Artiach and Arteche 2012), dust pollution (Reisen et al. 2014), river flow (Diongue and Ndongo 2016), electricity demand (Leschinski and Sibbertsen 2019), and traffic patterns (Ferrara and Guégan 2001). In economics and finance, the GARMA and k-factor GARMA models have been applied to study information related to equities (Beaumont and Smallwood 2019; Lu and Guegan 2011; Caporale and Gil-Alana 2014), interest rates (Ramachandran and Beaumont 2001; Gil-Alana 2007; Asai et al. 2018), inflation (Arteche and Robinson 2000; Caporale and Gil-Alana 2011; Peiris and Asai 2016), and unemployment (Gil-Alana 2007).

In spite of the intense interest in the GARMA model, there does not appear to be a unifying estimation approach for all model parameters. Several plausible estimators for λ, η, and ν exist, including parametric estimators in the time domain (Chung 1996a,b; Dissanayake et al. 2018) and frequency domain (Giraitis et al. 2001) and semi-parametric estimators (Hidalgo and Soulier 2004; Hidalgo 2005) extending the log-periodogrom estimators of Geweke and Porter-Hudak (1983) and Robinson (1995). However, a full set of accepted asymptotic results does not exist. Yajima (1996) showed that maximization of the periodogram could be used to consistently estimate the position of spectral poles, which would include ν. For semi-parametric estimators, Hidalgo and Soulier (2004) provide an asymptotic result for λ based on (1), demonstrating that the distribution for ν unknown is identical to that for ν known. Hidalgo (2005) extends these results and provides an estimator for ν that is asymptotically normal with rate of convergence T^{β}, with $\beta < 1$, and T denoting the sample size, whose distribution depends on whether $\nu = \{0, \pi\}$ or $\nu \in (0, \pi)$. Giraitis et al. (2001) provide the asymptotic distribution for λ for the parameterized Whittle estimator, and establish rate T convergence for the estimate of ν. Unfortunately, as discussed below, Giraitis et al. (2001) are unable to provide a full set of asymptotic results for their estimator of ν.

Obtaining valid inference for η is vital for researchers interested in obtaining confidence bands for estimated cycles and is imperative for those interested in tests for specific values, such as $\eta = 1$. Perhaps the most promising results were proposed by Chung (1996a,b) who argued that the constrained sum of squares (CSS) estimator for η converges at either rate T (if $|\eta| < 1$) or T^2 (if $|\eta| = 1$) to ratios of functionals of Brownian motion processes. Estimates for the remaining parameters achieve asymptotic normality at the standard rate of $T^{1/2}$. The CSS based results have recently been extended by Beaumont and Smallwood (2019) to consider multiple long memory cycles, while Peiris and Asai (2016) provide proposed distribution results for the estimator with heteroskedastic disturbances. Regrettably, given the potential discontinuity in the distribution for η coupled with the existence of a closed parameter space for this parameter, an initial consistency proof for the CSS estimators has proven quite elusive. Specifically, Chung and related extensions rely on the observation that the score evaluated at the true parameters is zero, which, as pointed out by Giraitis et al. (2001), may not be sufficient to establish consistency. Given the concerns regarding the theoretical results for the CSS estimator, it remains an open question as to whether or not the theory is practically useful.

In this paper, we review the properties of the two most commonly applied likelihood based estimators for the GARMA parameters, the CSS and Whittle estimators, and consider the consequences of using the distribution theory proposed by Chung (1996a,b). The results show that both estimators of η yield desirable results in terms of mean bias and root mean squared error. Further, the distribution theory of Chung is generally supported. Using an algorithm that does not bound the parameter space, however, there are important exceptions. For $|\eta| < 1$, it would appear that in finite samples, the estimates of η have slightly fatter tails and a more peaked density than theory implies. Although the problem may become negligible for very large samples, differences between theoretical and empirical densities can become severe for very small T and values of λ close to zero, such that asymptotic confidence bands can be quite unreliable. Further, the results show that the use of distribution theory under the case $|\eta| = 1$ can lead to severe size distortion potentially complicating issues for researchers attempting to test if a finite cycle exists.

Awaiting additional theoretical results, we can offer practical solutions to researchers interested in estimating confidence bands and employing statistical tests. For small estimated values of λ, researchers should avoid using asymptotic results when the sample size is small. Otherwise, results do show that narrower asymptotic confidence bands can be quite reliable, even for moderately small samples. Thus, a recommendation is made that when using theory, multiple sets of confidence intervals should be presented. Finally, in testing the hypothesis that $|\eta| = 1$, more conservative statistics based on the distribution theory under the alternative appear to yield more reliable results. A parametric bootstrap applied to US unemployment supports this assertion, and we can further recommend that whenever possible theoretical limitations can potentially be overcome through computational methods.

The rest of the paper is organized as follows. In Section 2, we motivate the problem using the US unemployment rate and present the two estimators of the GARMA process due to Chung (1996a,b) and Giraitis et al. (2001). In section 3, we present the Monte Carlo results, concentrating on the calculation of confidence bands. Section 4 offers evidence specific to testing$|\eta| = 1$, and a final section concludes.

2. Parametric Estimators of the GARMA Process

The two estimators analyzed here use time and frequency domain approximations for the Gaussian log likelihood function based on the GARMA process defined in (1). The first estimator we consider is a Whittle type estimator. For a sample size T, let $\tilde{T} = [T/2]$, where $[\cdot]$ denotes the integer part. Let Δ denote the set of all possible parameter values for $\delta = \{\phi', \theta', \lambda, \mu\}$ and let Q_T denote the set of Fourier frequencies, $\omega_j = 2\pi\, j/T$. Based on the spectrum $f(\omega_j)$ defined in (2), Giraitis et al. (2001) propose the following estimator (denoted GHR),

$$\begin{pmatrix} \hat{\delta} \\ \hat{\nu} \end{pmatrix} = \text{argmin}_{\Delta \times Q_T} S(\delta, \nu) \tag{3a}$$

where

$$S(\delta,\nu) = \left[\frac{1}{\tilde{T}}\sum_{j=0}^{T}\frac{I(\omega_j)}{f(\omega_j)}\right] \text{ and } I(\omega_j) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t e^{it\omega_j}\right|^2. \tag{3b}$$

Importantly, note that the estimate of ν is obtained with respect to the discrete set Q_T, and that the true value of ν need not be in this set. Under suitable regularity conditions, Giraitis et al. (2001) establish asymptotic normality for their estimate of δ, and prove that the estimate of ν is consistent. However, a limiting distribution for $\hat{\nu}$ is not available, since the function in (3) is not minimized for all values in the interval $[0,\pi]$.

Concentrating out the residual variance, and under a Gaussian assumption for ε_t, our second estimator is based on the constrained sum of squares (CSS) function of the parameters, $\phi',\ \theta',\ \lambda,\ \eta,\ \mu$ and is given by

$$L(\phi',\theta',\lambda,\eta,\mu) = -\frac{T}{2}\left[\ln(2\pi) + \ln\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\right) + 1\right]. \tag{4}$$

Under an initialization assumption, maximization of the CSS function produces a set of estimates that is asymptotically equivalent to the maximum likelihood values.

The proposed theory of Chung (1996b) establishes asymptotic normality for $\delta = (\lambda,\phi',\theta')'$, where, for example, the element from the information matrix for λ is given by $I_\lambda = 2(\pi^2/3 - \pi\nu + \nu^2)$. For the CSS estimate of η, which is found to be asymptotically independent of δ, Chung has the following proposed distributional result in his Theorem 2:

$$T(\hat{\eta}-\eta) \underset{d}{\rightarrow} \frac{\sin(\nu)}{\lambda}Y_0 = \frac{\sin(\nu)}{\lambda}\frac{\int_0^1 W_1 dW_2 - \int_0^1 W_2 dW_1}{\int_0^1 W_1^2(r)dr + \int_0^1 W_2^2(r)dr}, \quad \text{for } |\eta| < 1 \tag{5}$$

and

$$T^2(\hat{\eta}\pm 1) \underset{d}{\rightarrow} \mp\frac{1}{2\lambda}Y_1 = \mp\frac{1}{2\lambda}\frac{\int_0^1 \left[\int_0^r W(s)\,ds\right] dW(r)}{\int_0^1 \left[\int_0^r W(s)\,ds\right]^2 dr}, \quad \text{for } \eta = \pm 1 \tag{6}$$

where $W, W_1,$ and W_2 are independent Brownian motions. Percentiles of Y_0 and Y_1 can be simulated to yield confidence bands and test statistics for η.

To illustrate a concern regarding the use of these asymptotic results, the following GARMA (2,1) model was obtained using the CSS estimator for the seasonally adjusted US unemployment rate, u_t, from Jan. 1980 through July

2019,[1]

$$(1-2*0.9986L+L^2)^{0.326}(1-0.934L-0.026L^2)(u_t-6.16)=(1-0.567L)\varepsilon_t. \quad (7)$$

With $\hat{\eta}=0.9986<1$, the estimated model represents a stationary process, albeit highly persistent, with strong cycles. In contrast, if $\eta=1$, the model reduces to a non-stationary ARFIMA process lacking a finite cumulative impulse response function. The asymptotic 99% confidence bands constructed under the assumption $\eta=1$ are $[0.9980, 9987]$, while the corresponding quantities based on rate T convergence are $[0.9972, 1.0001]$.[2] Although more formal testing procedures advocated below generally yield evidence in support of $\eta<1$, casual observation of these confidence bands could prove confusing. The disparate conclusions also highlight one of the potential concerns as it relates to the use of the proposed distribution for the CSS estimator. In the following sections, we provide Monte Carlo analysis to shed light on the applicability of theoretical results both in forming confidence bands and for testing $\eta=1$.

3. The Monte Carlo Results

In this section, we present Monte Carlo results to assess how the estimators perform both in small and larger samples. From a computational perspective, an advantage of the Whittle based estimator of Giraitis et al. (2001) is its relative simplicity. For each Fourier frequency, ω_j, we minimize the function $S(\delta,\omega_j)$ in (3) with respect to δ, and track the value of the objective function for $j=0,1,\ldots,\tilde{T}$. The estimate of ν, $\hat{\nu}$, is the Fourier frequency associated with the minimum value of the objective function amongst the $\tilde{T}+1$ alternatives. Then, the estimate of δ is the value that minimizes the objective function with the frequency fixed at $\hat{\nu}$. An estimate of η can be obtained through the functional relationship, $\eta=\cos(\nu)$.

For the CSS estimator, note that the polynomial $(1-2\eta L+L^2)^{-\lambda}$ is related to the Gegenbauer polynomials, c_j, as follows (Gray et al. 1989):

$$(1-2\eta L+L^2)^{-\lambda}=\sum_{j=0}^{\infty}c_jL^j, \quad (8a)$$

[1] Data are from the St. Louis Federal Reserve Bank. The GARMA(2,1) model was selected on the basis of the AIC for all combinations of models with p and q less than 2.

[2] These quantities are calculated using (5) and (6) based on Chung's simulated values for Y_0 and Y_1. For example, the value of the 1st and 99th percentiles for Y_0 is -4.238 and 4.238. With $\nu=0.0525$ and a sample size of $T=475$, confidence bands can be calculated as $[-4.328\sin(\nu)/(475*0.326), 4.238\sin(\nu)/(475*0.326)]$.

where

$$c_j = \sum_{k=0}^{[j/2]} \frac{(-1)^k \Gamma(\lambda + j - k)(2\eta)^{j-2k}}{\Gamma(\lambda)\Gamma(k+1)\Gamma(j-2k+1)} \tag{8b}$$

and $\Gamma(\cdot)$ is the gamma function. Starting from $c_0 = 1$ and $c_1 = 2\eta\lambda$, the Gegenbauer polynomials can be computed recursively as

$$c_j = 2\eta\left(\frac{\lambda+1}{j} + 1\right)c_{j-1} - \left(2\frac{\lambda+1}{j} + 1\right)c_{j-2}. \tag{9}$$

The disturbance sequence in (1) is a function of the model parameters and in turn the Gegenbauer polynomials as follows,

$$\begin{aligned}\varepsilon_t = (1 - \phi_1 L - \cdots - \phi_p L^p)(x_t - \mu) - \sum_{j=1}^{t-1} c_j \varepsilon_{t-j} \\ - \theta_1 \sum_{j=0}^{t-2} c_j \varepsilon_{t-j-1} - \cdots - \theta_q \sum_{j=0}^{t-q-1} c_j \varepsilon_{t-j-q}.\end{aligned} \tag{10}$$

To calculate the model parameters, Chung (1996a,b) and Gray et al. (1989), advocate a line-search for η.[3] This implies that the parameter space being searched over is a discrete set. This seems especially unfortunate as the theory is developed under the assumption of a continuous parameter space and a potential advantage of the CSS estimator relative to the GHR counterpart is lost. In instances where the true value of η is unknown, a discretization implies that potentially large biases can result or that a very fine grid would need to employed. Further, if a large number of grid points is selected, it almost certainly becomes necessary to impose boundary constraints on η for computational purposes.

Here, we advocate the use of a double gradient-based procedure as in Ramachandran and Beaumont (2001). A set of starting values for η is selected, which is typically a grid from -1 to 1, using a relatively small step size to avoid potential local minima. Conditional on each value of η in the grid, an estimate of δ is obtained using a gradient based method. Conditional on the estimate of δ, we estimate the parameter η, again using a search algorithm. The procedure continues as we update the value of η along the grid. Once a neighborhood for the maximum value of the CSS function is obtained, we iterate using the two-step procedure until the norm of the estimated parameters between steps is sufficiently small. This procedure allows one to search over all possible values of η and does not impose boundary constraints that could artificially improve the fit of the CSS

[3] Chung (1996a,b) proposes estimation of δ using a gradient based method for each value η along the grid, while Gray et al. (1989) also estimate λ using a line search.

estimator in a simulation environment.

For the Monte Carlo experiments, we considered a total of eight different cases, including six GARMA(0,0) (Table 1) and two GARMA(1,0) models (Table 2). The true values of η are $\{-1, -0.9995, -0.50, 0.50, 0.9995, 1\}$ for the GARMA(0,0) cases. For $|\eta| = 1$, we fix $\lambda = 0.20$, where $\lambda = 0.40$ otherwise. For the GARMA(1,0) cases, we fixed $\lambda = 0.40$ and $\phi = 0.80$, allowing the true values of η to be 0.50 and 0.9995. Note that the last model has short memory dynamics and is parametrically close to the non-stationary border. We thus anticipate that this model may produce relatively poor results. For each model, we performed 2500 simulations and considered sample sizes of 100, 300, 500, 1000, and 2000 observations. To generate a data series, x_t, we calculated the autocovariances of the long memory processes and obtained the Cholesky factorization of the Toeplitz matrix.[4] This factorization is then multiplied by a sequence of normal random variates of the desired length. Data are generated through recursion for each GARMA(1,0) case, where μ is set to 0 throughout.

Tables 1-2 report the results of the mean bias and RMSE for each model and both the time domain estimator of Chung (CSS) and the frequency-based estimator of Giraitis et al. (2001) (GHR). To help interpret the results, the estimator that yields the smallest bias/RMSE in absolute value for a given sample size is shown in bold type. For both estimators, the absolute value of the mean bias associated with η is remarkably small, with a value that decreases rapidly with the sample size. The CSS outperforms the GHR estimator in terms of the mean bias of η. There are instances where the improvement in mean and RMSE can be somewhat large, especially when $|\eta| = 0.50$, likely resulting from the fact that the true value of η is not typically in the discrete parameter space for the GHR estimator (except when ν is 0 or π). When $\nu \neq 0$, the GHR estimator tends to dominate in mean bias for λ. In terms of RMSE, for $|\eta| \neq 1$, the CSS estimator tends to dominate for η, λ, and in the cases of the GARMA(1,0) model, for ϕ as well. It should be noted that the RMSE for both estimators of λ and ϕ are generally quite similar, and compare favorably with the computed asymptotic standard deviations of these parameters from Chung (1996b), Theorem 3, with one exception. In particular, performance of the estimators for the GARMA(1,0) model with $\eta = 0.9995$, $\lambda = 0.40$, and $\phi = 0.80$ tends to be quite poor. For sample sizes less than 2000, the CSS and GHR procedures can result in a mean bias for λ of -0.2023 and -0.3017, respectively. A similar picture emerges for ϕ, where the mean bias of ϕ can be as large as 0.0992 for the CSS estimator, while the GHR estimator tends to underestimate ϕ with a mean bias that is typically quite large in absolute value. The results for the GARMA(1,0) cases show that

[4]See McElroy and Holan (2012) for details concerning the autocovariances of Gegenbauer processes).

the mean bias in λ tends to be inversely related to the mean bias in ϕ , especially for the CSS estimator. As is well known to researchers using parametric estimators in the ARFIMA context, it can be difficult to distinguish high frequency components from low frequency pieces (Nielsen and Frederiksen 2005).

For researchers interested in obtaining point estimates for GARMA parameters, and η specifically, the GHR and CSS estimators appear to provide highly robust options. However, the question remains as to whether or not proposed distribution theory can be used for inference and the construction of confidence bands. To this end, Table 3 displays the estimated and theoretical percentiles from two potentially problematic cases, one with η near 1 and another with an autoregressive component.[5] Below the reported sample size, we present the percentiles of the distribution of the statistic calculated from Chung (1996a), using his equation 19 and Table 1, along with the empirical distribution of the same quantity resulting from both the CSS and GHR estimators.

We are primarily interested in the CSS estimator, noting two things regarding the GHR estimator. First, the empirical distribution of $T(\hat{\eta}-\eta)$ for the GHR estimator confirms the established convergence rate of T as shown by Giraitis et al. (2001). Second, we note that for the GHR estimator, the underlying parameter space is discrete. Consider for example, the empirical distribution of $T(\hat{\eta}-\eta)$ when the true values of η and λ are 0.9995 and 0.40 for a sample size equal to 300. Of the estimated 2500 values of η, 1162 are exactly equal to 0.99912, the closest possible value to 0.9995. While the estimator unquestionably performs well, this discretization can naturally result in small biases, which again helps to explain the findings in Tables 1-2, where the CSS estimator tends to dominate. It additionally highlights a potential concern in using a CSS-based algorithm that establishes a line search for η using a discrete set as in Chung (1996a) and Woodward et al. (1998). In what follows, we concentrate on the properties of the CSS estimator and algorithm proposed here.

For the two cases in Table 3, we generally see that the CSS estimator of η has an empirical distribution that is well approximated by the asymptotic distribution provided by Chung (1996a,b). The values of the empirical percentiles are typically quite close to the reported percentiles of Chung, especially for the 2.5% and 5.0% levels, which are important for statistical testing. Based on 500 observations, for example, with $\eta = 0.50$ and $\phi \neq 0$, the empirical 5^{th} percentile for $T(\hat{\eta}-\eta)$ is -4.41, which closely matches the proposed theoretical quantity equal to -4.70.[6]

[5] For brevity, we do not include all results from Tables 1 and 2, which are available upon request. Briefly summarizing, they provide conclusions that are qualitatively identical to those reported in Table 3 for the other parameterizations, except perhaps the case where $\phi = 0.80$ and $\eta = 0.9995$. Here, not surprisingly, for small samples, we find the distribution of is skewed left.

[6] We obtain similar results for the GARMA(1,0) case with $\eta = 0.50$, $\lambda = 0.40$, where for

In spite of generally confirming the proposed distributional results, Table 3 hints that the finite sample distribution may have a more peaked density and fatter tails than theory implies. Consider Figure 1, which provides kernel density plots of $2000(\hat{\eta} - \eta)$ and the corresponding theoretical quantity using the distribution theory for the CSS estimator based on the GARMA(0,0) model with $\eta = 0.50$ and $\lambda = 0.40$.[7] Indeed, the figure shows that there are several places where the associated kernel density plots cross, suggesting that in finite samples, the empirical distribution may have larger kurtosis and a more peaked density than implied by theory. We now turn to the question of whether these issues impact the practical usefulness of the distributional results in constructing confidence bands.

A large number of simulations generally reveal that theoretical confidence bands provide remarkably accurate coverage relative to empirical counterparts when $|\eta|$ is in the neighborhood of unity and/or λ is near the non-stationary boundary. However, there can be very serious concerns with the use of theory, especially when T is small and λ is near 0. Tables 4 and 5 highlight these issues, where we present the associated biases that would result from the use of theory in constructing confidence bands. More specifically, the tables report the difference between the theoretical value of η at the upper and lower 68%, 90%, 95%, and 99% confidence bands and the associated empirical quantity based on 5000 simulated values for GARMA(0,0) models with $\eta = 0.50$ and 0.98. For each value of η, we allow λ to take on the values $\{0.1, 0.2, 0.3, 0.4\}$, and as above, we consider several sample sizes ranging from 100 to 2000. As a reference, the theoretical bands for sample sizes of 500 observations are presented in bold font.

From Table 4, we see that the amplified empirical kurtosis is especially problematic for small T and λ. Generally speaking, the 99% bands appear to be uninformative when $\lambda = 0.10$, even for moderately large sample sizes. For $T = 500$, for example, 0.5% of estimated values of η are less than 0.2295, which starkly contrasts the theoretical value of 0.4799. Although somewhat reliable results can be obtained for 68% bands and sample sizes of at least 1000, the results with $\lambda = 0.10$ show that the existing theory may need to be exercised with some caution. We do note that the time series with $\lambda = 0.10$ might be viewed as somewhat extreme here, in light of the fact that they display characteristics that are difficult to distinguish from short memory. The theoretical first order autocorrelation coefficient, for example, is 0.1015, and after the first lag, there is no value in excess of 0.05 in absolute value.

example the empirical 5^{th} percentile $T(\hat{\eta} - \eta)$ is -4.11. The results support the proposed independence of η from other parameters.

[7] We use a Gaussian smoothing window and a bandwidth parameter of 3. To calculate the theoretical density, one needs the associated percentiles of Y_0 from (5). These values have been simulated using MATLAB code available on request, using equation 25 in Chung (1996a).

Remaining results generally support the use of CSS theory for construction of confidence bands, especially for samples larger than 100 and narrower bands. Overall, biases rapidly decrease with both sample size and the value of λ. With $T = 2000$ and $\lambda = 0.40$, as an example, 95% of all estimated values of η lie between 0.4962 and 0.5040, implying confidence bands for estimated cycles of between 5.975 and 6.0266 periods. These values are remarkably close to those implied by theory, where theoretical bands of 0.4970-0.5030 correspond to cycle lengths between 5.98 and 6.02 periods.

The results in Table 5 indicate that generally small biases in calculating confidence intervals with asymptotic quantities further decline as η approaches unity. Here for all cases, except when $\lambda = 0.10$ and T is smaller than 500, the empirical coverage areas are very well captured by asymptotic quantities. Especially when λ is large, the differences between empirical percentiles and the theoretical quantities becomes negligible. Displayed in the final panel in Table 5, we present results with $\lambda = 0.40$ where $\eta = 0.9995$, a parameterization very close to a unit root. Except in the case of the 99% confidence bands with $T = 100$, the associated biases are never greater than 0.0013 and are essentially zero for $T > 1000$.[8]

4. Hypothesis Testing for η

Overall, the results indicate that the proposed distribution theory for maximum likelihood-based estimators works well in constructing confidence bands as $|\eta| \to 1$. The question remains as to whether these results are useful for statistical testing purposes. Here, as recently emphasized by Dissanayake et al. (2018), there are several specific hypotheses of interest, including $H_0 : \eta = 1$ versus $H_A : \eta < 1$. Regrettably, as discussed above, a discontinuity exists in the proposed theory of Chung (1996a), such that we anticipate potential inferential problems. These concerns are validated in Table 6, which depicts the empirical distribution of the CSS estimator standardized by both T and T^2 when the true value of η is 1.

Turning to the specific percentile values, we note that the empirical distribution of $T^2(\hat{\eta} - 1)$ does not match the proposed asymptotic distribution of Chung (1996a). For example, in Table 6, the value of the empirical 1^{st} percentile for $T^2(\hat{\eta} - 1)$ can be more than 11 times larger than the value established by Chung.

[8] We also consider the case where $\eta = 0.9995$ and $\lambda = 0.10$. Relative to other cases with $\lambda = 0.10$, resulting biases are typically much smaller. For example, with $T = 100$ and 300, the theoretical lower bands of the 68% confidence intervals are 0.9958 and 0.9983, whereas the associated empirical quantities are 0.9750 and 0.9938. For all coverage areas, biases decrease rapidly with the sample size, even becoming as small as 0.0017 and 0.0002 for 99% lower and upper limits when T=2000.

In other words, the empirical distribution is dramatically more skewed left than the theoretical results would imply. Further, the empirical distribution takes on fewer positive entries than the proposed asymptotic distribution. For example, the value associated with the 99^{th} percentile from Chung's asymptotic distribution for $T^2(\hat{\eta} - 1)$ is 31.13. Based on a sample size of 300, this implies that when $\eta = 1$, 1% of all estimates of this parameter will be at least 1.00035. In contrast, the empirical distribution shows that only 1% of all values exceed 1.00018.

To analyze how empirical and theoretical distributional disparities impact inference, we consider the proposed tests of Chung (1996a,b) in Table 7 for $|\eta| = 1$, when the true value of η is 1 or -1. The hypothesis can be tested by constructing confidence intervals about the estimate of η, where an ARFIMA process cannot be rejected if the value of unity lies within the confidence interval. The left-hand side of Table 7 reports the empirical size based on the 95% and 99% confidence intervals when the true value of η is 1, while the right hand side of Table 7 presents the same results when $\eta = -1$. Note, the confidence intervals are constructed here using the distributional results with T^2 rate of convergence when $|\eta| = 1$ (see (6)). The table shows that the implementation of the proposed distribution theory under the null will result in massive size distortion, with only mild relief as the sample size increases. Consider the case where the generated data are ARFIMA processes with $\eta = 1$. Even with 2000 observations and a 5% test, the constructed theoretical confidence bands under H_0 fail to include unity 17.6% of the time. The rejection rates of the true null $\eta = 1$ can be larger than 20%. Moving to the 99% confidence intervals (e.g. a 1% test), we still see that the rejection rates exceed 13%. Throughout, the results are slightly worse when $\eta = -1$. For even large samples, these results suggest that the proposed distributional theory is unlikely to be useful to researchers interested in determining if the true data generating process is an ARFIMA or GARMA process. This can be especially problematic for those interested in testing for stationarity, where non-stationarity occurs for all values of $\lambda \geq 0.25$ when $|\eta| = 1$, but only occurs when $\lambda \geq 0.50$ otherwise. Clearly, suitable testing procedures are needed. We offer two possibilities that we wish to posit as potential avenues for future work.

A conservative approach in testing $|\eta| = 1$ would involve the use of both sets of confidence intervals that assume different rates of convergence. Recall from above that values of η in excess of unity are plausible, and thus the distribution theory for $|\eta| < 1$ is not applicable given a complex value for ν (see (5)). While it is quite clear that more work is needed in establishing an appropriate test, one possibility might be the implementation of a one-sided test for $\eta = 1$ versus the alternative $\eta < 1$, where any estimated value in excess of unity is taken as an automatic failure to reject. As an example, we implemented this approach for the generated ARFIMA data with the sample sizes ranging from 100 to 2000 observations using the test statistic from (5). The test statistic based on a 5%

test size that assumes rate T convergence, yielded rejection rates of 6.64%, 5.28%, 4.84%, 5.44%, and 4.36%, respectively.

A more formal test that is potentially preferred has recently been advocated by Dissanayake et al. (2018), who discuss the use of quasi-likelihood ratio test statistics based on a state space representation of the GARMA model. Following a similar approach, one could obtain the value of the likelihood function in (4), with and without $\eta = 1$ imposed, and form the test statistic as follows,

$$LR = 2\left[\max_{\phi',\theta',\lambda,\eta,\mu} L(\phi',\theta',\lambda,\eta,\mu) - \max_{\phi',\theta',\lambda,\mu} L(\phi',\theta',\lambda,1,\mu)\right]. \tag{11}$$

Under the null, the value of η lies on the boundary of the parameter space, and given the difficulties above, it seems likely that the distribution of the resulting test statistic will be non-standard. In light of the problems with the existing theory, as discussed here, computational methods may be preferred.

As a reasonable approach, one could use the estimated model under the null hypothesis to form a parametric bootstrap. Data of the desired length could be simulated based on sampling with replacement from the residuals of the null model, and the critical values of the distribution could be formed based on the test statistic in (11). This procedure was applied to the unemployment rate data described in Section 2, using the residuals from an estimated ARFIMA(2,1) model. Results, which are based on 5000 simulations, are presented in Table 7.[9] Here, we see that the likelihood ratio test statistic associated with the hypothesis $\eta = 1$ takes on a value of 6.31, exceeding simulated critical values at the 5%, but not at the 1% levels. As discussed above, these findings are likely of tremendous importance to applied researchers in economics, since the null model is non-stationary, but strong stationary cycles result under the alternative.

The results provide reasonably strong support favoring $|\eta| < 1$ and also highlight the difficulties that could be encountered for researchers employing CSS theory. More specifically, confidence bands constructed under the null hypothesis $|\eta| = 1$ suffer from such severe size distortion they are likely uninformative. For the example here, a more careful testing analysis shows a marginal rejection of $\eta = 1$, whereas strong rejection occurs for virtually any size when using the theory of Chung (1996a,b) based on the assumed T^2 rate of convergence. Finally, we see that the associated theoretical confidence bands under $|\eta| < 1$ only contain unity when considering 99% intervals, matching the likelihood ratio test results. In instances where a parametric bootstrap is not available, these findings suggest that more conservative bands may be informative in determining if long memory

[9]For computational purposes, and given the non-stationarity of the null model, the data are simulated recursively using the coefficients of the expansion of $(1-L)^{-0.3301}$ and the standard assumption that pre-sample observations are 0.

cycles have finite length.

5. Conclusions

Considerable attention has recently focused on the use of models that allow for long memory cycles and potential singularities in the spectral density function, notably Gegenbauer autoregressive moving (GARMA) and associated k-factor GARMA models. While a number of robust estimators appears to exist in the time and frequency domain, it does not appear that there are an accepted set of distributional results related to the parameter, or parameters in the case of k-factor models, dictating the cycle length and the positions of spectral poles.

In this manuscript, we analyze the performance of two relevant likelihood-based estimators, where at least partial distributional results have been proposed. This includes, the Whittle estimator (GHR) described by Giraitis et al. (2001), and the approximate maximum likelihood estimator (CSS) analyzed by Chung (1996a,b). While Giraitis et al. (2001) were able to prove rate-T convergence of the estimator of the pole, they were unable to provide an exact limiting distribution given its discrete nature. To date, only Chung (1996a,b) has proposed an approximate maximum likelihood estimator that could be used for statistical inference and the calculation on confidence bands. However, the results were obtained without a formal consistency proof, given complications in the distribution theory, including the fact that the relevant parameter space is closed and the proposed distribution appears to have a potential discontinuity. The proposed theory is potentially relevant for other likelihood-based estimators, serving at least as a reference point, and has recently been extended by Beaumont and Smallwood (2019) and Peiris and Asai (2016) to accommodate multiple poles and heteroskedastic disturbances. It is therefore imperative to understand the extent to which the proposed asymptotic theory is relevant.

Using a double gradient-based search algorithm, an extensive Monte Carlo analysis revealed that both estimators are highly robust in estimating model parameters, specifically the position of the spectral pole and its associated cosine, denoted η. As the CSS estimator admits a continuous parameter space, it is found to be relatively superior in estimation of η. Empirical results very strongly support the proposed distribution theory of Chung (1996a,b) for stationary parameterizations in a neighborhood of a unit root process, especially for larger samples. Additionally, in most cases, the distribution theory is shown to be quite useful in constructing confidence bands, where empirical coverage areas are sometimes surprisingly well captured by associated asymptotic quantities.

The findings are not without important caveats. For our estimator, in smaller samples, empirical distributions can have larger kurtosis than implied by theory. Although there may be evidence that this is a small sample problem, the simulation results show that constructed wider confidence bands are likely to be uninformative under very weak long memory in samples less than 500 observations. From a practical perspective, we would recommend the avoidance of the proposed asymptotic theory for very small samples when weak persistence is suspected, perhaps as evidenced by a differencing parameter in the neighborhood of zero. Otherwise, narrower confidence bands are generally found to be robust, such that we would also recommend using multiple sets of intervals when presenting theoretical results.

From a testing perspective, we further show that the proposed distribution theory under the null $|\eta| = 1$ would render severe size distortion and can cause confusion regarding whether a series is a stationary GARMA process or non-stationary ARFIMA/unit root process. This does not appear to be a small problem, with rejection rates under the null approaching 20% for a 5% test size in samples of about 2000. We are able to demonstrate however, that the more conservative use of theory under the assumption $|\eta| < 1$ may be useful in distinguishing ARFIMA/ARIMA processes from GARMA counterparts.

The paper concludes with an application to unemployment, where evidence using a parametric bootstrap provides support for the existence of stationary, long memory cycles in the labor market. The overall conclusions suggest that GARMA parameters can be well estimated by existing likelihood-based techniques. For researchers interested in obtaining confidence bands for cycle lengths and conducting inference, the existing theory also appears to be largely applicable. Nonetheless, the results also show there are still several gaps in the existing theory that merit additional exploration, specifically as it relates to testing for $|\eta|$ in the neighborhood of 1. Awaiting additional distributional results, we suspect computational methods are most likely to be informative for researchers interested in statistical inference.

References

Arteche, J. and Robinson, P. M. (2000). Semiparametric inference in seasonal and cyclical long memory processes. *Journal of Time Series Analysis*, 21(1):1–25.

Artiach, M. and Arteche, J. (2012). Doubly fractional models for dynamic heteroscedastic cycles. *Computational Statistics & Data Analysis*, 56(6):2139 – 2158.

Asai, M., Peiris, S., McAleer, M., and Allen, D. (2018). Cointegrated dynamics for a generalized long memory process: An application to interest rates. Technical Report EI2018-32, Econometric Institute Research Papers, University of Madrid.

Beaumont, P. M. and Smallwood, A. D. (2019). Constrained sum of squares estimation of multiple frequency long memory models. Working Paper, Florida State University.

Caporale, G. M. and Gil-Alana, L. (2014). Long-run and cyclical dynamics in the US stock market. *Journal of Forecasting*, 33(2):147–161.

Caporale, G. M. and Gil-Alana, L. A. (2011). Multi-factor Gegenbauer processes and European inflation rates. *Journal of Economic Integration*, pages 386–409.

Chung, C.-F. (1996a). Estimating a generalized long memory process. *Journal of Econometrics*, 73(1):237 – 259.

Chung, C.-F. (1996b). A generalized fractionally integrated autoregressive moving average process. *Journal of Time Series Analysis*, 17:111–140.

Diongue, A. K. and Ndongo, M. (2016). The k-factor GARMA process with infinite variance innovations. *Communications in Statistics-Simulation and Computation*, 45(2):420–437.

Dissanayake, G. S., Peiris, M. S., and Proietti, T. (2018). Fractionally differenced Gegenbauer processes with long memory: A review. *Statistical Science*, 33(3):413–426.

Ferrara, L. and Guégan, D. (2001). Forecasting with k-factor Gegenbauer processes: Theory and applications. *Journal of Forecasting*, 20(8):581–601.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238.

Gil-Alana, L. A. (2007). Testing the existence of multiple cycles in financial and economic time series. *Annals of Economics & Statistics*, 8(1):1–20.

Giraitis, L., Hidalgo, J., and Robinson, P. M. (2001). Gaussian estimation of parametric spectral density with unknown pole. *The Annals of Statistics*, 29(4):987–1023.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–24.

Gray, H. L., Zhang, N., and Woodward, W. A. (1989). On generalized fractional processes. *Journal of Time Series Analysis*, 10:233–257.

Hidalgo, J. (2005). Semiparametric estimation for stationary processes whose spectra have an unknown pole. *The Annals of Statistics*, 33(4):1843–1889.

Hidalgo, J. and Soulier, P. (2004). Estimation of the location and exponent of the spectral singularity of a long memory process. *Journal of Time Series Analysis*, 25(1):55–81.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165–76.

Leschinski, C. and Sibbertsen, P. (2019). Model order selection in periodic long memory models. *Econometrics and Statistics*, 9:78 – 94.

Lu, Z. and Guegan, D. (2011). Estimation of time-varying long memory parameter using wavelet method. *Communications in Statistics—Simulation and Computation*, 40(4):596–613.

McElroy, T. S. and Holan, S. H. (2012). On the computation of autocovariances for generalized Gegenbauer processes. *Statistica Sinica*, 22(4):1661–1687.

Nielsen, M. Ø. and Frederiksen, P. H. (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews*, 24(4):405–443.

Peiris, M. and Asai, M. (2016). Generalized fractional processes with long memory and time dependent volatility revisited. *Econometrics*, 4(4):37.

Ramachandran, R. and Beaumont, P. (2001). Robust estimation of GARMA model parameters with an application to cointegration among interest rates of industrialized countries. *Computational Economics*, 17(2/3):179–201.

Reisen, V. A., Zamprogno, B., Palma, W., and Arteche, J. (2014). A semiparametric approach to estimate two seasonal fractional parameters in the SARFIMA model. *Mathematics and Computers in Simulation*, 98:1 – 17.

Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23(3):1048–1072.

Woodward, W. A., Cheng, Q. C., and Gray, H. L. (1998). A k-factor GARMA long-memory model. *Journal of Time Series Analysis*, 19(485-504).

Yajima, Y. (1996). Estimation of the frequency of unbounded spectral densities. ASA Proceedings Business and Economic Statistics Section 4-7, American Statistical Association.

Table 1: Bias and RMSE for Whittle and CSS estimates of GARMA(0,0) model parameters

T	CSS η Bias	GHR η Bias	CSS λ Bias	GHR λ Bias	CSS η RMSE	GHR η RMSE	CSS λ RMSE	GHR λ RMSE	Chung Asymp
Model #1: GARMA(0,0); $\eta = 1$, $\lambda = 0.20$									
100	**-0.00644**	-0.00659	**0.00564**	-0.03109	0.0262	**0.0237**	**0.0444**	0.0591	0.0390
300	**-0.00089**	-0.00093	**0.00311**	-0.01106	0.0058	**0.0043**	**0.0251**	0.0278	0.0225
500	**-0.00027**	-0.00028	**0.00228**	-0.00630	0.0017	**0.0012**	**0.0194**	0.0199	0.0174
1000	**-0.00006**	-0.00008	**0.00173**	-0.00320	**0.0003**	0.0004	**0.0131**	0.0133	0.0123
2000	**-0.00002**	-0.00002	0.00210	**-0.00158**	**0.0001**	0.0002	0.0095	**0.0092**	0.0087
Model #2: GARMA(0,0); $\eta = -1$, $\lambda = 0.20$									
100	0.00724	**0.00624**	**0.00498**	-0.03160	0.0277	**0.0248**	**0.0448**	0.0593	0.0390
300	**0.00090**	0.00091	**0.00306**	-0.01025	0.0052	**0.0039**	**0.0252**	0.0276	0.0225
500	0.00031	**0.00027**	**0.00249**	-0.00670	0.0019	**0.0012**	**0.0191**	0.0205	0.0174
1000	**0.00007**	0.00009	**0.00183**	-0.00342	**0.0004**	0.0004	**0.0134**	0.0138	0.0123
2000	0.00002	**0.00002**	**0.00167**	-0.00176	**0.0002**	0.0002	0.00935	**0.0092**	0.0087
Model #3: GARMA(0,0); $\eta = 0.9995$, $\lambda = 0.40$									
100	**-0.00070**	-0.00118	0.01696	**0.00227**	**0.0029**	0.0044	**0.0470**	0.0546	0.0396
300	**-0.00008**	-0.00018	0.01057	**0.00999**	**0.0005**	0.0008	**0.0274**	0.0316	0.0229
500	**-0.00002**	-0.00006	**0.00809**	0.00918	**0.0002**	0.0004	**0.0213**	0.0245	0.0177
1000	**1.18E-06**	-1.99E-06	0.00562	**-0.00107**	**0.0001**	0.0001	0.0149	**0.0133**	0.0125
2000	**1.59E-06**	2.22E-06	0.00495	**-0.00007**	**0.0001**	0.0001	0.0108	**0.0092**	0.0089
Model #4: GARMA(0,0); $\eta = -0.9995$, $\lambda = 0.40$									
100	**0.00089**	0.00120	0.01566	**0.00180**	**0.0039**	0.0042	**0.0453**	0.0458	0.0390
300	**0.00009**	0.00014	0.01049	**0.00929**	**0.0006**	0.0007	**0.0276**	0.0309	0.0229
500	**0.00002**	0.00006	**0.00823**	0.00839	**0.0003**	0.0004	**0.0214**	0.0243	0.0177
1000	1.63E-06	**1.43E-06**	0.00553	**-0.00123**	**0.0001**	0.0001	0.0149	**0.0131**	0.0125
2000	6.80E-07	**-3.76E-06**	0.00452	**-0.00022**	**0.0001**	0.0001	0.0106	**0.0093**	0.0089
Model #5: GARMA(0,0); $\eta = 0.50$, $\lambda = 0.40$									
100	**0.00102**	-0.00587	0.02535	**0.00602**	**0.0298**	0.0437	**0.0805**	0.1050	0.0675
300	**0.00029**	0.00051	**0.01620**	-0.01930	**0.0102**	0.0114	**0.0458**	0.0480	0.0390
500	**0.00001**	0.00152	0.01245	**0.00467**	**0.0060**	0.0079	**0.0361**	0.0387	0.0302
1000	**-0.00002**	-0.00080	0.00754	**0.00472**	**0.0030**	0.0038	**0.0248**	0.0268	0.0214
2000	**-0.00002**	0.00036	0.00521	**0.00397**	**0.0016**	0.0020	**0.0170**	0.0630	0.0151
Model #5: GARMA(0,0); $\eta = -0.50$, $\lambda = 0.40$									
100	**-0.00002**	0.00621	0.02950	**-0.02176**	**0.0313**	0.0429	**0.0812**	0.1039	0.0675
300	**-0.00006**	-0.00016	**0.01535**	-0.01945	**0.0102**	0.0113	**0.0447**	0.0490	0.0390
500	**0.00002**	-0.00151	0.01112	**0.00245**	**0.0063**	0.0080	**0.0346**	0.0374	0.0302
1000	**-0.00005**	0.00063	0.00736	**0.00456**	**0.0031**	0.0039	**0.0239**	0.0260	0.0214
2000	**-0.00001**	-0.00035	0.00460	**0.00328**	**0.0016**	0.0020	0.0166	**0.0020**	0.0151

Notes: The asymptotic standard error for λ using the theory of Chung (1996a) is presented in the last column. Quantities appearing in bold font indicate the smaller bias/RMSE between the CSS and GHR estimators.

Table 2: Bias and RMSE for Whittle/CSS based estimates of the GARMA(1,0) model

T	CSS: η	GHR: η	CSS: λ	CSS: λ	CSS: ϕ	GHR: ϕ	Chung
Model #7: GARMA(1,0); $\eta = 0.50$, $\lambda = 0.40$, $\phi = 0.80$							(Asymp λ)
			Parameter Bias				
100	**-0.00058**	-0.00901	**0.01937**	-0.05005	-0.02735	**-0.02068**	0.06779
300	-0.00036	**-0.00005**	**0.01320**	-0.02692	-0.01035	**-0.00357**	0.03914
500	**-0.00016**	0.00119	0.00999	**-0.00106**	**-0.00683**	-0.00724	0.03032
1000	**-0.00001**	-0.00086	0.00710	**0.00249**	-0.00348	**-0.00319**	0.02144
2000	**-0.00001**	0.00037	0.00655	**0.00488**	**-0.00146**	-0.00150	0.01516
Model #7: GARMA(1,0); $\eta = 0.50$, $\lambda = 0.40$, $\phi = 0.80$							(Asymp ϕ)
			Parameter RMSE				
100	**0.03235**	0.04947	**0.07735**	0.11326	**0.07617**	0.07828	0.06024
300	**0.01075**	0.01152	**0.04481**	0.05360	0.03854	**0.03769**	0.03478
500	**0.00641**	0.00784	**0.03435**	0.03796	**0.02918**	0.03093	0.02694
1000	**0.00334**	0.00408	**0.02422**	0.02584	**0.01951**	0.02004	0.01905
2000	**0.00161**	0.00201	**0.01736**	0.01867	**0.01369**	0.01394	0.01347
T	CSS: η	GHR: η	CSS: λ	CSS: λ	CSS: ϕ	GHR: ϕ	Chung
Model #8: GARMA(1,0); $\eta = 0.9995$, $\lambda = 0.40$, $\phi = 0.80$							(Asymp λ)
			Parameter Bias				
100	**-0.01416**	-0.02782	**-0.20227**	-0.30174	**0.09923**	0.10871	0.12635
300	**-0.00223**	-0.00796	-0.15377	**-0.11110**	**0.09992**	-0.11546	0.07295
500	**-0.00065**	-0.00337	-0.12132	**-0.01833**	**0.07958**	-0.24442	0.05651
1000	**-0.00009**	-0.00066	-0.08719	**0.00541**	**0.06242**	-0.13533	0.03996
2000	**0.000002**	-0.00004	**0.01609**	0.01929	**-0.02224**	-0.10542	0.02825
Model #8: GARMA(1,0); $\eta = 0.9995$, $\lambda = 0.40$, $\phi = 0.80$							(Asymp ϕ)
			Parameter RMSE				
100	**0.03787**	0.05880	**0.24362**	0.33197	0.22420	**0.20178**	0.06024
300	**0.00981**	0.01753	**0.19453**	0.24230	**0.15666**	0.41871	0.03478
500	**0.00466**	0.00822	**0.16658**	0.20506	**0.14172**	0.48856	0.02694
1000	**0.00162**	0.00255	**0.12891**	0.14197	**0.12002**	0.33196	0.01905
2000	**0.00006**	0.00037	**0.04347**	0.10770	**0.07160**	0.27240	0.01347

Notes: The asymptotic standard error for λ using the theory of Chung (1996a) is presented in the first panel for both models and in the last column. The analogous quantity for ϕ is presented in the second panel, and again in the last column. Quantities appearing in bold font indicate a smaller bias/RMSE between the CSS and GHR estimators.

Table 3: Empirical distribution for the percentiles $T(\hat{\eta}-\eta)$

Model: GARMA(0,0) with $\eta = 0.9995$, $\lambda = 0.40$										
	0.001	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
Chung	**-0.335**	**-0.284**	**-0.220**	**-0.172**	**-0.124**	**0.124**	**0.172**	**0.220**	**0.284**	**0.335**
CSS/100	-1.639	-1.268	-0.689	-0.386	-0.169	0.047	0.067	0.092	0.120	0.140
GHR/100	-3.092	-1.721	-0.739	-0.739	-0.147	0.050	0.050	0.050	0.050	0.050
CSS/300	-1.002	-0.628	-0.319	-0.194	-0.096	0.066	0.115	0.149	0.166	0.182
GHR/300	-1.493	-0.902	-0.442	-0.442	-0.113	0.084	0.150	0.150	0.150	0.150
CSS/500	-0.722	-0.431	-0.275	-0.155	-0.075	0.071	0.137	0.202	0.247	0.260
GHR/500	-0.737	-0.737	-0.382	-0.382	-0.105	0.092	0.211	0.211	0.250	0.250
CSS/1000	-0.460	-0.368	-0.209	-0.122	-0.072	0.079	0.152	0.229	0.345	0.494
GHR/1000	-0.763	-0.467	-0.211	-0.211	0.007	0.007	0.184	0.184	0.322	0.421
CSS/2000	-0.369	-0.276	-0.189	-0.118	-0.064	0.064	0.126	0.236	0.343	0.461
GHR/2000	-0.421	-0.421	-0.194	-0.194	-0.194	0.013	0.201	0.201	0.368	0.516
Model: GARMA(1,0) with $\eta = 0.50$, $\lambda = 0.40$, $\phi = 0.80$										
	0.001	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
Chung	**-9.176**	**-7.786**	**-6.012**	**-4.700**	**-3.399**	**3.399**	**4.700**	6**6.012**	**7.786**	**9.176**
CSS/100	-14.111	-10.745	-6.949	-4.284	-2.348	2.226	4.476	7.491	10.681	12.771
GHR/100	-19.098	-13.188	-7.422	-7.422	-1.825	3.583	3.583	8.779	13.7421	22.897
CSS/300	-13.187	-11.189	-7.236	-4.840	-2.388	2.048	4.125	6.488	10.263	12.863
GHR/300	-11.011	-11.011	-5.474	-5.474	0.000	0.000	5.408	5.408	10.748	16.018
CSS/500	-14.513	-10.683	-6.761	-4.417	-2.164	1.975	3.966	6.564	10.153	13.467
GHR/500	-14.648	-14.648	-9.123	-3.636	-3.636	1.812	7.220	7.220	12.587	12.587
CSS/1000	-12.619	-9.842	-6.535	-3.972	-2.139	2.107	4.117	6.774	9.783	13.259
GHR/1000	-12.750	-12.750	-7.273	-7.273	-1.815	3.623	3.623	9.041	9.041	14.440
CSS/2000	-15.543	-11.177	-6.657	-3.872	-2.046	2.043	3.752	6.869	10.212	13.488
GHR/2000	-14.545	-9.083	-9.083	-3.630	-3.630	1.813	7.246	7.246	12.670	12.670

Notes: In bold font, we present the values at theoretical percentiles for the test statistic using equation (4) and the associated simulated quantities for Y_0 from Chung (1996a). The remaining elements yield the associated values at a given percentile for $\hat{\eta}$ using the CSS and GHR estimators.

Table 4: Bias in estimating confidence bands in GARMA(0,0) models with $\eta = 0.05$.

	68%L	68%U	90%L	90%U	95%L	95%U	99%L	99%U
Theory: 500	**0.4799**	**0.5201**	**0.4623**	**0.5376**	**0.4517**	**0.5482**	**0.4267**	**0.5733**
BIAS WITH $\lambda = 0.10$								
CSS/100	0.2504	-0.1563	0.4764	-0.2420	0.6080	-0.2297	0.8701	-0.1318
CSS/300	0.0518	-0.0627	0.1154	-0.1123	0.1753	-0.1429	0.4368	-0.2568
CSS/500	0.0328	-0.0361	-0.0641	-0.0657	0.0830	-0.0766	0.1623	-0.1394
CSS/1000	0.0140	-0.0178	-0.0290	-0.0317	0.0354	-0.0378	0.0469	-0.0499
CSS/2000	0.0077	-0.0090	-0.0151	-0.0158	0.0175	-0.0194	0.0231	-0.0273
Theory: 500	**0.4900**	**0.5101**	**0.4812**	**0.5188**	**0.4759**	**0.5241**	**0.4634**	**0.5366**
BIAS WITH $\lambda = 0.20$								
CSS/100	0.0110	-0.0251	0.0858	-0.1001	0.1279	-0.1321	0.2952	-0.2209
CSS/300	0.0045	-0.0092	-0.0303	-0.0336	0.0435	-0.0435	0.0655	-0.0674
CSS/500	0.0032	-0.0056	0.0166	-0.0201	0.0239	-0.0248	0.0348	-0.0396
CSS/1000	0.0026	-0.0029	0.0098	-0.0093	0.0122	-0.0122	0.0198	-0.0174
CSS/2000	0.0016	-0.0018	0.0055	-0.0053	0.0068	-0.0067	0.0105	-0.0085
Theory: 500	**0.4933**	**0.5067**	**0.4874**	**0.5125**	**0.4839**	**0.5161**	**0.4756**	**0.5244**
BIAS WITH $\lambda = 0.30$								
CSS/100	-0.0161	0.0162	0.0044	-0.0095	0.0200	-0.0326	0.0553	-0.0731
CSS/300	-0.0042	0.0041	0.0050	-0.0046	-0.0115	-0.0118	0.0256	-0.0214
CSS/500	-0.0019	0.0017	0.0019	-0.0040	0.0064	-0.0080	0.0157	-0.0168
CSS/1000	-0.0006	0.0005	0.0022	-0.0020	0.0040	-0.0040	0.0081	-0.0077
CSS/2000	0.0001	-0.0001	0.0018	-0.0018	0.0027	-0.0028	0.0043	-0.0042
Theory: 500	**0.4950**	**0.5050**	**0.4906**	**0.5094**	**0.4879**	**0.5120**	**0.4817**	**0.5183**
BIAS WITH $\lambda = 0.40$								
CSS/100	-0.0181	0.0195	-0.0301	0.0305	-0.0266	0.0269	-0.0151	-0.0064
CSS/300	-0.0054	-0.0057	-0.0076	0.0075	-0.0053	0.0040	0.0037	-0.0034
CSS/500	-0.0030	0.0031	-0.0036	0.0028	-0.0016	0.0013	0.0022	-0.0022
CSS/1000	-0.0012	0.0012	-0.0006	0.0007	0.0004	-0.0001	0.0027	-0.0011
CSS/2000	-0.0003	0.0002	0.0004	-0.0006	0.0008	-0.0010	0.0022	-0.0023

Notes: The table reports the difference between the value of η associated with theoretical confidence bands, which have been constructed using equation (5) along with simulated values for Y_0, and the estimated value of η associated with a given percentile. 68%L and 68%U refer to 68% lower and upper confidence bands, with similar meaning for other quantities. Values appearing in bold font are the actual values of η (e.g. not the biases) that are generated from equation (5) for a sample size of 500 observations.

Table 5: Bias in estimating confidence bands in GARMA(0,0) models with $\eta = 0.98$.

	68%L	68%U	90%L	90%U	95%L	95%U	99%L	99%U
Theory: 500	**0.9754**	**0.9846**	**0.9713**	**0.9866**	**0.9689**	**0.9911**	**0.9632**	**0.9968**
			BIAS WITH $\lambda = 0.10$					
CSS/100	0.1104	0.0063	0.9097	0.0145	0.9732	-0.0211	1.1108	0.0267
CSS/300	0.0151	-0.0059	0.0336	-0.0052	0.0478	-0.0015	0.0987	0.0078
CSS/500	0.0079	-0.0066	0.0172	-0.0089	0.0218	-0.0079	0.0358	-0.0031
CSS/1000	0.0036	-0.0036	0.0073	-0.0062	0.0099	-0.0068	-0.0158	-0.0071
CSS/2000	0.0018	-0.0018	-0.0036	-0.0032	0.0043	-0.0037	0.0061	-0.0051
Theory: 500	**0.9777**	**0.9823**	**0.9757**	**0.9843**	**0.9745**	**0.9855**	**0.9716**	**0.9884**
			BIAS WITH $\lambda = 0.20$					
CSS/100	0.0105	0.001	0.0047	0.0036	0.0760	0.0081	0.2348	0.0206
CSS/300	0.0023	-0.0014	0.0090	-0.0049	0.0129	-0.0057	0.0235	-0.0047
CSS/500	0.0013	-0.0010	0.0055	-0.0037	0.0067	-0.0045	0.0120	-0.0051
CSS/1000	0.0007	-0.0006	0.0025	-0.0020	0.0035	-0.0027	0.0056	-0.0035
CSS/2000	0.0004	-0.0004	0.0013	-0.0012	0.0017	-0.0016	0.0027	-0.0021
Theory: 500	**0.9785**	**0.9815**	**0.9771**	**0.9829**	**0.9763**	**0.9837**	**0.9744**	**0.9856**
			BIAS WITH $\lambda = 0.30$					
CSS/100	-0.0030	0.0041	0.0047	0.0028	0.0107	0.0020	0.0387	0.0081
CSS/300	-0.0009	0.0009	0.0015	-0.0007	0.0036	-0.0020	0.0071	-0.0028
CSS/500	-0.0004	0.0005	0.0012	-0.0008	0.0024	-0.0015	0.0047	-0.0030
CSS/1000	-0.0001	0.0001	0.0007	-0.0005	0.0011	-0.0011	0.0021	-0.0019
CSS/2000	0.00002	-0.00002	0.0005	-0.0004	0.0007	-0.0006	0.0013	-0.0012
Theory: 500	**0.9788**	**0.9812**	**0.9778**	**0.9822**	**0.9772**	**0.9828**	**0.9758**	**0.9842**
			BIAS WITH $\lambda = 0.40$					
CSS/100	-0.0040	0.0043	-0.0068	0.0072	-0.0064	0.0070	0.0034	0.0059
CSS/300	-0.0012	0.0013	-0.0016	0.0018	-0.0007	0.0010	0.0005	-0.0005
CSS/500	-0.0007	0.0007	-0.0006	0.0009	-0.0001	0.0005	0.0004	-0.0002
CSS/1000	-0.0003	0.0003	-0.0001	0.0003	0.0001	0.0001	0.0009	-0.0005
CSS/2000	-0.0001	0.0001	0.0001	-0.0001	0.0002	-0.0002	0.0007	-0.0004
			GARMA(0,0) model $\eta = 0.9995$, $\lambda = 0.40$					
Theory: 500	**0.9900**	**0.9997**	**0.9992**	**0.9998**	**0.9991**	**0.9999**	**0.9998**	**1.0002**
			BIAS FOR MODEL NEAR UNIT ROOT BOUNDARY					
CSS/100	-0.0006	0.0004	-0.0008	0.0009	0.0001	0.0011	0.0066	0.0017
CSS/300	-0.0002	0.0002	-0.0003	0.0002	0.00002	0.0003	0.0013	0.0005
CSS/500	-0.0001	0.0001	-0.0001	0.0001	0.00004	0.0001	0.0002	0.0002
CSS/1000	-0.00005	0.00003	-0.00004	-0.00003	0.00001	-0.00003	0.0001	-0.00004
CSS/2000	-0.00002	0.000001	0.00001	-0.00003	0.00003	-0.00004	0.00009	-0.00006

Notes: The table reports the difference between the value of η associated with theoretical confidence bands, which have been constructed using equation (5) along with simulated values for Y_0, and the estimated value of η associated with a given percentile. 68%L and 68%U refer to 68% lower and upper confidence bands, with similar meaning for other quantities. Values appearing in bold font are the actual values of η (e.g. not the biases) that are generated from equation (5) for a sample size of 500 observations.

Table 6: Percentiles for $T(\hat{\eta}-\eta)$ and $T^2(\hat{\eta}-\eta)$ for the empirical distribution of η when $|\eta|=1$.

MODEL: GARMA(0,0): $\eta = 1.00$, $\lambda = 0.20$										
$T(\hat{\eta}-\eta)$	0.001	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990	0.995
Chung	**-244.44**	**-185.29**	**-121.23**	**-81.00**	**-75.13**	**10.14**	**15.44**	**21.53**	**31.18**	**39.85**
CSS:T=100										
$T(\hat{\eta}-\eta)$	-17.791	-13.262	-7.194	-3.825	-1.238	0.074	0.101	0.133	0.167	0.211
$T^2(\hat{\eta}-\eta)$	-1779.1	-1326.2	-719.39	-382.49	-123.84	7.365	10.106	13.297	16.654	21.114
CSS:T=300										
$T(\hat{\eta}-\eta)$	-9.974	-6.843	-2.328	-0.846	-0.297	0.024	0.032	0.042	0.053	0.060
$T^2(\hat{\eta}-\eta)$	-2992.3	-2052.8	-698.42	-253.80	-89.105	7.306	9.543	12.469	15.824	18.050
CSS:T=500										
$T(\hat{\eta}-\eta)$	-5.020	-3.332	-1.116	-0.494	-0.1774	0.014	0.019	0.024	0.031	0.037
$T^2(\hat{\eta}-\eta)$	-2510.1	-1666.2	-557.87	-246.83	-88.696	7.041	9.350	11.796	15.253	18.704
CSS:T=1000										
$T(\hat{\eta}-\eta)$	-2.072	-1.4230	-0.593	-0.286	-0.087	0.007	0.009	0.012	0.015	0.017
$T^2(\hat{\eta}-\eta)$	-2072.1	-1429.8	-592.49	-286.06	-86.90	6.839	9.208	11.978	14.823	17.289
CSS:T=2000										
$T(\hat{\eta}-\eta)$	-0.911	-0.644	-0.247	-0.010	-0.034	0.003	0.005	0.006	0.007	0.009
$T^2(\hat{\eta}-\eta)$	-1821.5	-1288.0	-493.68	-199.01	-68.111	6.656	8.930	11.097	14.235	18.194

Notes: In bold font, we present the values at theoretical percentiles for the test statistic $T^2(\hat{\eta}-\eta)$ using equation (6) and the associated simulated quantities for Y_1 from Chung (1996a). The remaining elements yield the associated values at a given percentile for $\hat{\eta}$ using the CSS estimator.

Table 7: Rejection rates of the null hypothesis $|\eta| = 1$ using Chung's confidence intervals

	GARMA(0,0) $\eta = 1,\ \lambda = 0.20$		**GARMA(0,0)** $\eta = -1,\ \lambda = 0.20$	
Sample Size	99%	95%	99%	95%
100	0.1780	0.2184	0.1808	0.2268
300	0.1468	0.1864	0.1704	0.2236
500	0.1492	0.1848	0.1724	0.2180
1000	0.1460	0.1812	0.1528	0.1964
2000	0.1328	0.1756	0.1424	0.1844

Notes: The results use the estimated value of λ based on equation (6) to calculate 95% and 99% confidence bands about the estimated value of η when $|\eta| = 1$, using the CSS estimator. Quantities here indicate the proportion of occurrences where the associated upper bands are less than 1, for $\eta = 1$, or where the lower band is greater than -1, when the true value of $\eta = -1$.

Table 8: Results of parametric bootstrap applied to US unemployment

	η	λ	ϕ_1	ϕ_2	θ	SSR	CSS
ARFIMA:	1	0.3301	0.9234	0.0292	-0.5481	11.613	207.421
GARMA:	0.9986	0.3263	0.9341	0.0263	-0.5674	11.459	210.574
Confidence bands for $\hat{\eta}$ under the assumption $\lvert\eta\rvert < 1$							
	90%L	90%U	95%L	95%U	99%L	99%U	
	0.9979	0.9994	0.9977	0.9995	0.9972	1.0001	
Bootstrapped test statistic for $H_0 : \eta = 1$							
Test stat: $2(CSS_u - CSS_r) = 6.3067(0.0224)$							
Bootstrapped Critical Values							
		1%	5%	10%			
		3.7671	5.0200	7.8154			

Notes: The ARFIMA model has been obtained with the restriction $\eta = 1$ imposed. Given the functional relationship $d = 2\lambda$, the model can be written as:

$$(1 - L)^{0.6602}(1 - 0.923L - 0.029L^2)(u_t - 6.20) = (1 - 0.548L)\varepsilon_t.$$

CSS denotes the value of the likelihood function in (4) with and without the restriction $\eta = 1$ imposed. The test statistic has been calculated as in equation (11). Critical values have been calculated on the basis of a parametric bootstrap, where the residuals from the ARFIMA model above have been randomly sampled with replacement to generate 5000 samples of 475 observations using the parameters of the ARFIMA model. For computational purposes, the set of starting values for η range from 0.925 to 1, where no constraints have been imposed on the parameter space. Using the estimated values of the CSS function with and without the constraints imposed, the test statistic in (11) has been calculated and the associated critical values are presented in the final panel.

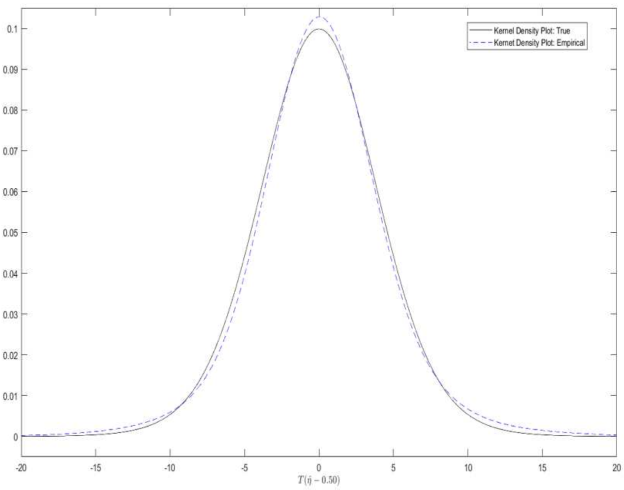

Figure 1: Theoretical and Empirical Kernel Density Plot of $T(\hat{\eta} - \eta)$ for $T = 2000$ and $\eta = 0.50$.